INFORMATION AND CONTROL 25, 145–165 (1974)

File Organization: Consecutive Storage of Relevant Records on Drum-Type Storage

SAKTI P. GHOSH

IBM Research Laboratory, San Jose, California 95114

Certain structure relationships between a query set, a record set, and storage media provide an opportunity to organize the record set without redundancy on the storage media, in such a manner that all records pertinent to any query in the query set can be retrieved with minimum access time. This property between query sets and record sets has been studied for drum-type storage media. Sufficient conditions for such a property have been established. It has been shown that the two dimensional storage capability of the drum-type storage can be utilized to extend the class of query sets and record sets for which the consecutive retrieval property exists on a linear storage media.

1. INTRODUCTION

The inverted file organization has been advocated and used in practice for some years. The most common type of inversion used occurs when a file is inverted with respect to the values of only one of its fields (i.e., attributes). The most important advantage of this type of inverted file is that all the records in which the particular attribute takes a fixed value can be stored in adjacent storage locations with minimum access delays, and thus the retrieval time is reduced considerably. In a *simple* formatted file, an attribute can take only one value in a record and hence redundant storage of records will not occur. Thus, the *simple* formatted inverted file has associated with it the concept of minimum retrieval time utilizing minimum storage space. This property of a file organization has been defined as the *consecutive retrieval* property (C-R property) by the author (1970).

As users of computer information become more sophisticated, their queries become more complex. The queries involve specifying multiple values of multiple attributes with complex logical relationships between these specifications. The complexities of both the query structure and file structure leads to the possibility that a record may be pertinent to more than one query or to a collection of values of multiple attributes. Thus, if the most important advantage of an inverted file (i.e., minimum access time for relevant records), is to be achieved for such a situation, redundant storage of records becomes necessary in many circumstances. The necessity of redundant storage of records depends on the structure of the file and the query.

In dealing with file organization problems, the structure of the storage device becomes important. Such a device may be considered as a collection of storage segments, say S_i. In each segment one and only one record can be stored. Associated with this collection of storage segments $\{S_i\}$ is an access time matrix $((t(S_i, S_j)))$, where $t(S_i, S_j)$ is the access time from S_i to S_j. In some cases $t(S_i, S_j)$ may also include the time needed to read the information contained in S_j. For many storage devices $t(S_i, S_j)$ may not be a symmetric function because the time needed to access S_j from S_i may not be the same as the access time from S_j to S_i. In a linear access storage; i.e., one-dimensional (1D) device; e.g., tape, track of a disk or drum, etc., the access time from S_i to S_j varies directly as the linear distance between S_i and S_j. Thus, on an one dimensional storage device the Consecutive Retrieval property is equivalent to consecutive storage of relevant records. In previous work (1970) the author has discussed the existence of the C-R property on one dimensional storage.

With the advent of drum- and disk-type storage, a second dimension has been added to storage devices. In these devices a set of recording heads, one (and in some new types of storage more than one) per recording track, are mounted on a vertical stand. In a drum storage the vertical stand remains fixed, whereas, in a disk package storage, the vertical stand can move and, thus, generate a hypothetical drum for each fixed position of the stand. At any instant in time, all the recording heads are in a position to read or record information from its associated recording track but the control unit permits only one head to be active. Hence, these storage devices will be referred to as two-dimensional (2D) storage with continual displacement of the recording head in one dimension. (Some new storage devices can activate simultaneously two recording heads but, in such situations, the two heads are reading or writing segments of the same record to increase the throughput. This additional capability does not change the results discussed in this paper.) In discussing the C-R property on 2D storage with continual displacement of the recording heads in 1D, it is assumed that the records are of equal length. This assumption is needed to maintain synchronism between the records on different tracks and, thus, eliminate access delays between them. The array of records which are stored along any track are referred to as the *primary array* of the 2D organization. The direction of the primary array is referred to as primary direction. The array of records which are accessible to the set of recording heads at any instant of time are referred to as the *secondary array* of the 2D organization and this direction is referred to as the secondary direction.

The consecutive retrieval property of a query set with respect to a record set on a 2D storage with continual displacement of the recording head in 1D may be defined as follows.

DEFINITION. A set of queries is said to have 2D C-R property w.r.t. a set of records, if the records can be stored in a two dimensional storage without duplication, such that the records pertinent to any query are stored in consecutive storage locations in the primary direction, though not necessarily in the same primary array.

In the preceding definition the continual displacement of the recording head in 1D in the 2D storage has not been stated because in this paper only drum-type storage devices are discussed and this particular property is associated with them. It should also be noted that, for the existence of the 2D C-R property, the organized records need not form a complete lattice. In this paper, an attempt is made to show that the 2D organization capability of the storage device can be utilized to achieve consecutive retrieval in organizing records. It is also shown that the class of file structures and query structures for which the C-R property exists on 1D storage can be extended when 2D storage is used.

The usefulness of a 2D C-R organization can be illustrated by an application to air-flight schedules. Suppose the following air-flight schedule is provided and it is necessary to query in the form "list all flights to..." and "list all flights from..."

Flight #	Departing city	Destination city
UA 001	SF	NY
AA 002	LA	CH
NWA 003	KA	CH
UA 004	SF	CH
TWA 005	SE	NY
UA 006	LA	NY
UA 007	SF	WA
TWA 008	LA	WA
AA 009	KA	NY
NWA 010	KA	WA
AA 011	SE	WA
UA 012	SE	CH

In order to form a 2D C-R organization, flights departing from the same city are to be stored on the same track. The secondary direct is used to store flights having the same destination. As the reading heads can be read only one record at a time from each secondary array, hence diagonal columns are used to store flights having the same destination city. Thus the following type of organization is appropriate.

Departing cities ↓ | Destination cities ⟶

SF	NY	WA	CH			
LA		NY	WA	CH		
KA			NY	WA	CH	
SE				NY	WA	CH

This arrangement can be converted into a complete latice as follows:

Departing cities ↓ | Destination cities ⟶

SF	NY	WA	CH
LA	CH	NY	WA
KA	WA	CH	NY
SE	NY	WA	CH

Thus the 2D C-R organization for the air-flight schedule is as follows:

Tracks			
→	UA 001	UA 007	UA 004
→	AA 002	UA 006	TWA 008
→	NWA 010	NWA 003	AA 009
→	TWA 005	AA 011	UA 012

⟶
Direction of rotation

This 2D C-R organization can also be used to answer queries which are subsets of the original two queries.

2. Mathematical Formulation

Suppose a query set $\{Q\}$ contains the queries $Q_1, Q_2, \ldots, Q_n$ and the record set $\{R\}$ contains the records $R_1, R_2, \ldots, R_m$. The set of records which are pertinent to a query Q_i is denoted by $\rho(Q_i)$. In some very special situations, when $\rho(Q_i)$ for $i = 1, 2, \ldots, n$ are all disjoint sets; e.g. if the Q_i's represent keys of records, etc., then the 2D C-R property exists between $\{Q\}$ and $\{R\}$. However, in many practical situations, all the $\rho(Q_i)$ are not disjoint.

Let $\rho(Q_i) = \{R_{i1}, R_{i2}, \ldots R_{im_i}\} \subseteq \{R\}$. Suppose the records are stored in a 2D organization and the storage location of the jth record pertinent to the ith query be denoted by $(\alpha_{ij}, \beta_{ij})$, where α_{ij} denotes the primary array and β_{ij} the secondary array. Thus, the queries and storage locations of pertinent records may be represented by the following two-way representation:

$$\begin{array}{cccc}
Q_1 & Q_2 & \cdots & Q_n \\
(\alpha_{11}, \beta_{11}) & (\alpha_{21}, \beta_{21}) & & (\alpha_{n1}, \beta_{n1}) \\
(\alpha_{12}, \beta_{12}) & (\alpha_{22}, \beta_{22}) & & (\alpha_{n2}, \beta_{n2}) \\
(\alpha_{1m_1}, \beta_{1m_1}) & (\alpha_{2m_2}, \beta_{2m_2}) & & (\alpha_{nm_n}, \beta_{nm_n})
\end{array} \tag{2.1}$$

If there exists a 2D organization of all the records of $\{R\}$ without duplication such that in every column of β's the β_{ij}'s are a set of consecutive integers then $\{Q\}$ has the 2D C-R property w.r.t. $\{R\}$.

Definition. The incidence domain of a record is defined to be the set of queries for which the record is pertinent.

Thus, the incidence domain of R_i is denoted by

$$\rho^{-1}(R_i) = \{Q_j \in \{Q\} \mid R_i \in \rho(Q_j)\}.$$

The same definition of incidence domain may be extended to a set of records; thus, the incidence domain of a set of records may be denoted by

$$\rho^{-1}\{R\} = \bigcup_{R_i \in \{R\}} \{Q_j \in \{Q\} \mid R_i \in \rho(Q_j)\}.$$

Thus, if two sets of records have the same incidence domains then they will be defined as record sets with *equivalent incidence domains.*

Two records R_i and R_j are said to have disjoint incidence domains when $\rho^{-1}(R_i) \cap \rho^{-1}(R_j) = \varnothing$, where $\varnothing$ is the empty set. Disjoint incidence domains play an important role in 2D C-R organization. In a 2D C-R organization the records in a secondary array must have disjoint incidence domains or they will destroy the 2D C-R organization.

Given any set of records $\{R\}$ and a query set $\{Q\}$ it is always possible to partition $\{R\}$ into a number of clusters say $\{R_{1i}\}$ $\{R_{2i}\}, \ldots, \{R_{ki}\}$ such that the records belonging to any cluster have disjoint incidence domains w.r.t. $\{Q\}$. In general, clusters with disjoint incidence domains within them can be formed in more than one manner. Given the record clusters it is possible to form the Record–Cluster query incidence matrix.

$$\begin{array}{c} \\ \{R_{1i}\} \\ \{R_{2i}\} \\ \{R_{3i}\} \\ \vdots \\ \{R_{ki}\} \end{array} \begin{array}{c} \begin{array}{cccc} Q_1 & Q_2 & Q_3 \cdots & Q_n \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ & & & \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}. \tag{2.2}$$

The $(i - j)$th position of the matrix contains a 1 if Q_j contains a pertinent record in $\{R_{il}\}$, and 0 otherwise. As $\{R_{il}\}$ contains records which have disjoint incidence domains hence it cannot contain more than one record pertinent to Q_j. The 2D C-R property between $\{Q\}$ and $\{R\}$ may be stated in terms of properties of the matrix (2.2). If there exists a partition of $\{R\}$ into cluster, which have disjoint incidence domains w.r.t. $\{Q\}$ within them, and there exists at least one permutation of these clusters for which the matrix (2.2) has consecutive 1's in each column, then there exists the 2D C-R property between $\{Q\}$ and $\{R\}$. If there exists no such partition with this property then the 2D C-R property does not exist between $\{Q\}$ and $\{R\}$.

In the following section some sufficient conditions between $\{Q\}$ and $\{R\}$ for the existence of the 2D C-R property are discussed.

3. The Two-Dimensional Consecutive Retrieval Property

In previous work by the author (1970) sufficient conditions for the existence of the 1D C-R property between a query set and a record set have been discussed, Suppose the 1D C-R property exists between query set $\{Q_1\}$ and record set $\{R_1\}$, then the 1D C-R organization of $\{R_1\}$ can be considered as a primary array of a 2D organization. Similarly if the 1D C-R property exists between $\{Q_2\}$ and $\{R_2\}$ then a 1D C-R organization of $\{R_2\}$ can be considered as another primary array of the 2D organization. In order that this 2D organization be a 2D C-R organization it is necessary that each secondary array have disjoint incidence domain. If $\{R_1\}$ and $\{R_2\}$ have disjoint incidence domain then the 2D organization is a 2D C-R organization. This process can be iterated, hence the following lemma is evident.

LEMMA 1. *If the 1D C-R property exists between $\{Q_i\}$ and $\{R_i\}$ for $i = 1, 2,..., k$ and the $\{R_i\}$'s for $i = 1, 2,..., k$ have disjoint incidence domains then the 2D C-R property exists between $\bigcup_{i=1}^{k} \{Q_i\}$ and $\bigcup_{i=1}^{k} \{R_i\}$.*

It should be noted that the conditions of Lemma 1 are also a sufficient condition for existence of the 1D C-R property between $\bigcup_{i=1}^{k} \{Q_i\}$ and $\bigcup_{i=1}^{k} \{R_i\}$. The practical utility of Lemma 1 is that if a drum-type storage is available and no 1D type storage is available then the records $\bigcup_{i=1}^{k} \{R_i\}$ can be organized on the drum as a 2D C-R organization.

In many practical situations it may be difficult to find 1D C-R organizations with disjoint incidence domains. The following two theorems provide some sufficient conditions under which two 1D C-R organizations can be combined into a 2D C-R organization.

THEOREM 1. *If $\{Q_{ij}\}$ has the 1D C-R property w.r.t. $\{R_{ij}\}$, $i = 1, 2$ and the following conditions are satisfied.*

(i) *If $\{R_{i_1 j}\}$ contains a record which is pertinent to a query in $\{Q_{i_2 j}\}$ (where $i_1 \neq i_2$ and $i_1, i_2 = 1, 2$) then that record is also contained in $\{R_{i_2 j}\}$.*

(ii) *If there exists at least one 1D C-R organization of $\{R_{1j}\}$ w.r.t. $\{Q_{1j}\}$ and one 1D C-R organization of $\{R_{2j}\}$ w.r.t. $\{Q_{2j}\}$ in which $\{R_{1j}\} \cap \{R_{2j}\}$ have the same sequential ordering and are stored in consecutive storage locations, then $\{Q_{1j}\} \cup \{Q_{2j}\}$ has the 2D C-R property w.r.t. $\{R_{1j}\} \cup \{R_{2j}\}$.*

The proof of the theorem is given in the appendix.

THEOREM 2. *If $\{Q_{ij}\}$ has the 1D C-R property w.r.t. $\{R_{ij}\}$, $i = 1, 2$ and the following conditions are satisfied.*

(i) *If $\{R_{i_1 j}\}$ contains a record which is pertinent to a query in $\{Q_{i_2 j}\}$ (where $i_1 \neq i_2$ and $i_1, i_2 = 1, 2$) then that record is also contained in $\{R_{i_2 j}\}$.*

(ii) *If there exists a 1D C-R organization of $\{R_{1j}\}$ for the query set $\{Q_{1j}\}$ and an 1D C-R organization of $\{R_{2j}\}$ for the query set $\{Q_{2j}\}$ in which, the* *elements of* $\{R_{1j}\} \cap \{R_{2j}\}$ *have the same sequential ordering and the number of records between any two elements of* $\{R_{1j}\} \cap \{R_{2j}\}$ *in the two* 1*D C-R organizations are the same, then* $\{Q_{1j}\} \cup \{Q_{2j}\}$ *have the* 2*D C-R property w.r.t.* $\{R_{1j}\} \cup \{R_{2j}\}$.

The proof of the theorem is given in the appendix.

DEFINITION. A set of records is defined to be a cover for a set of queries if the incidence domains of the records, w.r.t. the query set, are disjoint and the union of the incidence domains is equal to the query set.

Sets of records which are covers (also referred to as cover sets) play an important role in 2D C-R organization. Suppose $\{R\}$ is a record set and it can be partitioned into k subsets $\{R_1\}, \{R_2\},..., \{R_k\}$ where each subset is a cover set for the query set $\{Q\}$. If a 2D organization is formed in which the cover sets $\{R_i\}$, $i = 1, 2,..., k$ are stored in consecutive secondary arrays, then the 2D organization is also a 2D C-R organization w.r.t. $\{Q\}$, because every secondary array contains one and only one record pertinent to a query in $\{Q\}$ and every query in $\{Q\}$ contains a pertinent record in every secondary array. This result is summarized in the following lemma.

LEMMA 2. *If a set of records can be partitioned into cover sets w.r.t. a query set, then the query set has the* 2*D C-R property w.r.t. the record set.*

The cover sets can also be used for augmenting a 2D C-R organization. Suppose a query set $\{Q\}$ has the 2D C-R property w.r.t. a record set $\{R\}$ and a 2D C-R organization of $\{R\}$ contains a secondary array, say $\{R_c\}$, which is a cover of $\{Q\}$. Then another cover, say $\{R_1\}$, which is disjoint from $\{R\}$ can be stored as a secondary array adjacent to $\{R_c\}$ without destroying the 2D C-R organization. Thus, $\{Q\}$ has the 2D C-R property w.r.t. $\{R\} \cup \{R_1\}$. This leads to the following lemma.

LEMMA 3. *The* 2*D C-R property of a query set w.r.t. a record set is invariant under the addition of disjoint cover sets to the record set provided the original record set contains a cover set as a secondary array in a* 2*D C-R organization.*

EXAMPLE 1. Consider a formatted file with binary valued attributes. For simplicity it is assumed that there are three attributes A_1, A_2, and A_3. Each attribute can take the values 0 or 1. Thus, there are eight possible records which may be represented as 111, 110, 101, 011, 100, 010, 001, 000. This set of records is also referred to as the set of all binary records over three attributes.

The query set consists of queries which can specify the presence of an attribute; i.e., $A_i = 1$, or can specify no value for the attribute; i.e., $A_i = X$. Thus, there are eight queries which are of the form XXX, $XX1$, $X1X$, $1XX$, $X11$, $1X1$, $11X$, and 111. This set of queries is referred to as the set of all binary queries over three attributes.

It has been shown by the author (1970) that the three queries $XX1$, $X1X$, and $1XX$ do not have 1D C-R property w.r.t. the set of all binary records. The record 000 is pertinent to none of the queries, hence, it may be omitted. The other seven records can be grouped into four subsets such that each is a cover for the query set $\{XX1, X1X, 1XX\}$.

These covers are

$$\begin{Bmatrix}101\\010\end{Bmatrix} \begin{Bmatrix}011\\100\end{Bmatrix}, \begin{Bmatrix}001\\110\end{Bmatrix}, \{111\}.$$

Using these cover sets as secondary arrays, a 2D C-R organization may be constructed as follows:

$$\begin{pmatrix}101, 011, 001, 111\\010, 100, 110 \quad\end{pmatrix}.$$

The retrieval scheme for the three queries, showing the coordinates of storage locations of pertinent records are given by the following table.

Queries →	$XX1$	$X1X$	$1XX$
Storage locations	(1, 1)	(2, 1)	(1, 1)
of pertinent	(1, 2)	(1, 2)	(2, 2)
records →	(1, 3)	(2, 3)	(2, 3)
	(1, 4)	(1, 4)	(1, 4)

The second coordinates of the storage locations of the records pertinent to each query are consecutive integers, hence, the 2D organization is a 2D C-R organization.

In constructing the 2D C-R organization it is important to find sufficient conditions related to adding queries to query sets, under wich the 2D C-R organization is invariant. The following theorem is an effective tool for extending 2D C-R organizations.

THEOREM 3. *The 2D C-R property of a query set is invariant under the addition of a new query if a 2D C-R organization does not contain more than* *one pertinent record of the new query in any secondary array and the total number of pertinent records of the new query is greater than the maximum number of records in the primary direction of the 2D C-R organization.*

The proof of this theorem is given in the appendix. An obvious corollary to this theorem is as follows.

COROLLARY 3.1. *The union of a set of queries with 1D C-R property and another query has the 2D C-R property provided the number of records pertinent to the new query is greater than or equal to the number of records pertinent to the query set.*

In spite of the flexibility achieved by using 2D storage over 1D storage it is not true that any set of three queries has the 2D C-R property. However, the following theorem provides some sufficiently mild conditions under which the 2D C-R property exists for three queries.

THEOREM 4. *A sufficient condition for the existence of the 2D C-R property between Q_1, Q_2, Q_3 and its pertinent records is the negation of two or more of the following conditions:*

$$|\rho(Q_1Q_2\bar{Q}_3)| > |\rho(\bar{Q}_1\bar{Q}_2Q_3)|; \ |\rho(Q_1\bar{Q}_2Q_3)| > |\rho(\bar{Q}_1Q_2\bar{Q}_3)|;$$
$$|\rho(\bar{Q}_1Q_2Q_3)| > |\rho(Q_1\bar{Q}_2\bar{Q}_3)|;$$

where $|\rho(Q_1Q_2\bar{Q}_3)|$ *is the number of records which are pertinent to* Q_1 *and* Q_2 *but not* Q_3, *etc.*

The proof of this theorem is given in the appendix.

Records with disjoint incidence domains play an important role in the 2D C-R organization. As stated previously, they form secondary arrays in a 2D C-R organization. They also can be used to determine invariance of a 2D C-R organization under augmenting of new records. Suppose $\{r_1\}, \{r_2\},..., \{r_n\}$ are the secondary arrays of a 2D C-R organization of $\{R\}$ for the query set $\{Q\}$. Let $\{R_1\}$ be another disjoint (w.r.t. $\{R\}$) record set with disjoint incidence domain (w.r.t. $\{Q\}$) within it and has an equivalent incidence domain with one of the secondary arrays, say $\{r_i\}$. If $\{R_1\}$ is stored as a secondary array between $\{r_i\}$ and $\{r_{i+1}\}$ or between $\{r_{i-1}\}$ and $\{r_i\}$, the 2D C-R property of the 2D organization remains invariant. This type of construction can be extended to additional sets of records which satisfy conditions similar to $\{R_1\}$. Hence, the following theorem is true.

THEOREM 5. *The 2D C-R property of a query set w.r.t. a record set is invariant under the addition of another disjoint set of records to the record set, provided the new record set can be partitioned into subsets where each subset has*

(i) *disjoint incidence domains within it, and*

(ii) *equivalent incidence domain with a secondary array of a fixed 2D C-R organization of the old set of records.*

EXAMPLE 2. Consider a formatted file with three attributes A_1, A_2, and A_3. A_1 can take three values v_{11}, v_{12}, and v_{13}. A_2 can take three values v_{21}, v_{22}, and v_{23}. A_3 can take four values v_{31}, v_{32}, v_{33}, and v_{34}. The records are denoted by $R_k = (v_{1i_1}, v_{2i_2}, v_{3i_3})$ where $k = i_3 + (i_2 - 1)4 + (i_1 - 1)12$. The 2D C-R property exists between the query set $\{xv_{21}v_{32}, xxv_{32}, v_{11}xx\}$ and the record set $\{R_1, R_7, R_{10}, R_{14}, R_{26}\}$. A 2D C-R organization is given by

$$\begin{array}{c} R_{26}, R_{14}, R_{10}, R_7 . \\ R_1 \end{array}$$

Consider another set of records $\{R_2, R_6\}$. This set of record is disjoint from $\{R_1, R_7, R_{10}, R_{14}, R_{26}\}$. The incidence domain of R_2 is $\{xv_{21}v_{32}, xxv_{32}, v_{11}xx\}$, which is equivalent to the incidence domain of the secondary array $\{R_{14}, R_1\}$. Hence R_2 can be stored to the right or left of this secondary array without destroying the 2D C-R organization. The incidence domain of R_6 is $\{xxv_{32}, v_{11}xx\}$ which is equivalent to the incidence domain of R_{10}. Hence, the following 2D organization is a 2D C-R organization for the query set $\{xv_{21}v_{32}, xxv_{32}, v_{11}xx\}$.

$$\begin{array}{c} R_{26}, R_2, R_{14}, R_6, R_{10}, R_7 . \\ R_1 , \end{array}$$

Monotonicity (in a set theoretic sense) among incidence domains of subsets of records can be used to identify the 2D C-R property between a query set and a record set. The following theorem provides such an opportunity.

THEOREM 6. *A sufficient condition for the existence of the 2D C-R property between a query set and a record set is that the records can be partitioned into subsets with disjoint incidence domains within each subset and one of the following conditions are satisfied:*

(i) *The incidence domains of the subsets form a monotone increasing or decreasing sequence, or*

(ii) *The incidence domains of the subsets form monotone decreasing sequences on both sides of a maximum element.*

The proof of this theorem is given in the appendix.

Relations between incidence domains of two adjacent secondary arrays of a 2D C-R organization may be used as a test for deciding invariance of the 2D C-R organization for augmenting new records to the record set. The following theorem provides an opportunity for such situations.

THEOREM 7. *The 2D C-R property of a query set w.r.t. a record set is invariant under the addition of a disjoint set of records to the record set, provided the incidence domain of the new record set satisfies the following conditions:*

(i) *The records belonging to new record set have disjoint incidence domains which are subsets of the query set, and*

(ii) *The incidence domain of the new record set contains the intersection and is contained in the union, of the incidence domains of two adjacent secondary arrays in a 2D C-R organization of the old record set.*

The proof of the theorem is given in the appendix.

A special case of Theorem 7 is the following lemma.

LEMMA 5. *Any query set has the 2D C-R property w.r.t. a record set provided the record set can be partitioned into two subsets such that the records belonging to each subset have disjoint incidence domains among themselves.*

The results of Theorem 7 can easily be generalized and the generalization is given in the following lemma.

LEMMA 6. *The 2D C-R property of a query set w.r.t. a record set is invariant under the additon of another disjoint record set to the record set provided it can be partitioned into subsets which satisfy the condition of Theorem 7.*

The proof of this lemma follows as a mathematical induction of Theorem 7.

EXAMPLE 3. Consider the structure of the file discussed in Example 2. Suppose the query set given is $\{1xx, 2xx, 12x, 13x, 2x3, x2x\}$ and the record set is $\{R_2, R_7, R_8, R_{12}, R_{15}, R_{23}\}$. Then a 2D C-R organization is of the following form:

$$\begin{array}{llll} R_2, & R_8, & R_7, & R_{12} \\ & & R_{23}, & R_{15}. \end{array}$$

Suppose another set of record $\{R_3, R_6, R_{19}\}$ is added to the record set. This record set can be divided into subsets $\{R_6\}$ and $\{R_3, R_{19}\}$ which have disjoint incidence domain within them. The incidence domain of $\{R_8\}$ is $\{1xx, 12x, x2x\}$ and the incidence domain of $\{R_7, R_{23}\}$ is $\{1xx, 12x, x2x, 2xx, 2x3\}$. The incidence domain of $\{R_6\}$ is $\{1xx, 12x, x2x\}$ which is contained in the union and contains the intersection of the incidence domain of $\{R_8\}$ and $\{R_7, R_{23}\}$. Thus, $\{R_6\}$ can be inserted between them as a secondary array without destroying the 2D C-R organization. Similarly the incidence domain of $\{R_3, R_{19}\}$ is $\{1xx, 2xx, 2x3\}$ which is contained in the union and contains the intersection of the incidence domain of $\{R_7, R_{23}\}$ and $\{R_{12}, R_{15}\}$, hence, the subset $\{R_3, R_{19}\}$ can be inserted as a secondary array without destroying the 2D C-R organization.

Thus, the following 2D organization will be a 2D C-R organization for the query set.

$$\begin{array}{llllll} R_2, & R_8, & R_6, & R_7, & R_3, & R_{12} \\ & & & R_{23}, & R_{19}, & R_{15}. \end{array}$$

4. Discussion

All file organizations which are based on rigid structures, designed to reduce access time and/or redundant storage, have the disadvantage that they are vulnerable to updating. 2D C-R organization also shares the same disadvantage. It is difficult to give any analytic expressions for the vulnerability because it depends on the nature of updating. Though some preliminary work has been done on updating yet a global theory is lacking and would be a good topic for research.

The 2D C-R property when applicable to a query set and record set could lead to some unoccupied spaces in the organization. These unoccupied spaces have the same effect on space requirement as redundant storage. It is interesting to note that unlike 2D C-R organizations the 1D C-R organizations do not have unoccupied spaces but the class of query sets and record sets for which 2D C-R organizations is applicable, contain the class (of query sets and record sets) for which 2D C-R organization is applicable. This is consistent with the basic theory of file organizations, i.e., "Access time reduction is obtained at the cost of space redundancy and vice versa."

The amount of unoccupied spaces in the 2D C-R organization will depend on the query set and the record set. Some of the theorems, given in the paper, discuss situations when there will be no unoccupied spaces in the 2D C-R organization. A general theorem linking unoccupied spaces in a 2D C-R organization to the structure of the query set and the record set would be difficult to obtain. The technique of subdividing the query set and obtaining a 2D C-R organization between each subset of query set and the record set separately, can be used to eliminate the unoccupied spaces, but it is the feeling of the author that in most situations such attempts will result in more redundancy.

Appendix

Proof of Theorem 1. Suppose a 1D C-R organization of $\{R_{1j}\}$ for the query set $\{Q_{1j}\}$ satisfying condition (ii) is

$$R_{11}, R_{12}, ..., R_{1n_1}.$$

Similarly, a 1D C-R organization of $\{R_{2j}\}$ for the query set $\{Q_{2j}\}$ satisfying condition (ii) may be represented as

$$R_{21}, R_{22}, ..., R_{2n_2}.$$

From condition (ii) $\{R_{1j}\} \cap \{R_{2j}\}$ must be of the form

$$\{R_{1j}\} \cap \{R_{2j}\} = \{R_{1j_1} = R_{2l}, R_{1j_1+1} = R_{2l+1}, ..., R_{1j_1+k} = R_{2l+k}\}$$

for some values of j_1, l, and k.

Using condition (i) the following 2D organization can be constructed, which has the C-R property for the query set $(Q_{1j}\} \cup \{Q_{2j}\}$

$$\begin{array}{llllllll} R_{11}, R_{12}, ..., & R_{1j_1-1}, & R_{1j_1}, & R_{1j_1+1}, & ..., & R_{1j_1+k}, & R_{1j_1+k+1}, & ..., R_{1n_1}. \\ R_{21}, ..., & R_{2l-1}, & & & & & , R_{2l+k+1} & ..., R_{2n_2}. \end{array}$$

In the above organization the 1D C-R organization of both $\{R_{1j}\}$ and $\{R_{2j}\}$ are unaltered in the primary direction and no two records of the 1D C-R organization are in the same secondary array.

This completes the proof.

Proof of Theorem 2. Suppose the two 1D C-R organizations satisfying condition (ii) are denoted by

$$R_{11}, R_{12}, ..., R_{1n_1}$$

and

$$R_{21}, R_{22}, ..., R_{2n_2}.$$

By virtue of condition (ii) the elements of $\{R_{1j}\} \cap \{R_{2j}\}$ are of the following form:

$$\{R_{1j}\} \cap \{R_{2j}\} = \{R_{1j_1} = R_{2k_1}, R_{1j_1+l_1} = R_{2k_1+l_1}, \ldots, R_{1j_1+l_p} = R_{2k_1+l_p}\}$$

for some j_1, k_1, l_1,..., l_p where $0 < l_1 < l_2 < \cdots < l_p$.

According to condition (i) there cannot exist any record pertinent to $\{Q_{1j}\}$ in $\{R_{2j}\}$ other than the elements of $\{R_{1j}\} \cap \{R_{2j}\}$ and the same is true for $\{Q_{2j}\}$ and $\{R_{1j}\}$.

Hence, a 2D C-R organization of $\{R_{1j}\} \cup \{R_{2j}\}$ for the query set $\{Q_{1j}\}\{Q_{2j}\}$ can be constructed as follows:

$$\begin{array}{llllll}
 & R_{11}, \ldots, R_{1j_1-1}, & R_{1j_1}, & R_{1j_1+1}, \ldots, & R_{1j_1+l_1}, & R_{1j_1+l_1+1}, \ldots \\
R_{21}, R_{22}, \ldots, R_{2k_1-1}, & & & , R_{2k_1+1}, \ldots, & & , R_{2k_1+l_1+1}, \ldots \\
 & & & , R_{1j_1+l_p}, & R_{1j_1+l_p+1}, \ldots, R_{1n_1} & \\
 & & & , & , R_{2k_1+l_p+1}, \ldots, R_{2n_2} &
\end{array}$$

In this 2D organization any secondary array corresponding to the elements of $\{R_{1j}\} \cap \{R_{2j}\}$ will have only one record whereas the secondary arrays between any two elements of $\{R_{1j}\} \cap \{R_{2j}\}$ will contain two records when disjoint incidence domains. If there are any secondary arrays to the left of R_{1j_1} or the right of $R_{1j_1+l_p}$ they will also have disjoint incidence domains.

This completes the proof.

Proof of Theorem 3. Suppose the query set $\{Q\}$ has the 2D C-R property w.r.t. $\{R\}$, and the maximum number of records in the primary direction of a 2D C-R organization is denoted by m. Q_1 is another query with $|\rho(Q_1)| \geqslant m$. According to the assumption in the theorem, in the particular 2D C-R organization, no secondary array contains more than one element of $\rho(Q_1)$.

Suppose there are m_1 positions in the primary direction which contain a record pertinent to Q_1. In each of the other $m - m_1$ primary positions a record belonging to $\rho(Q_1)$ but not present in the 2D C-R organization can be stored in an available storage location in the secondary array. Thus, Q_1 will have 2D C-R organization w.r.t. these m records. The remaining $|\rho(Q_1)| - m$ records of $\rho(Q_1)$ can be stored in as many consecutive storage locations in the primary direction adjacent to the 2D C-R organization. Thus, $\{Q\} \cup \{Q_1\}$ has the 2D C-R property.

This completes the proof.

Proof of Theorem 4. The proof of this theorem is based on three lemmas. They are as follows.

LEMMA 4.1. *A sufficient condition for three queries* $\{Q_1, Q_2, Q_3\}$ *to have the 2D C-R property is the negation of either or both of the following two conditions*:

$$|\rho(\bar{Q}_1Q_2Q_3)| > |\rho(Q_1\bar{Q}_2\bar{Q}_3)|; \; |\rho(Q_1\bar{Q}_2Q_3)| > |\rho(\bar{Q}_1Q_2\bar{Q}_3)|.$$

Proof of Lemma 4.1. The records pertinent to the three queries can be partitioned into the following seven sets. $\rho(Q_1\bar{Q}_2\bar{Q}_3)$, $\rho(\bar{Q}_1Q_2\bar{Q}_3)$, $\rho(\bar{Q}_1\bar{Q}_2Q_3)$, $\rho(Q_1Q_2\bar{Q}_3)$, $\rho(Q_1\bar{Q}_2Q_3)$, $\rho(\bar{Q}_1Q_2Q_3)$, $\rho(Q_1Q_2Q_3)$.

Case 1. If $|\rho(\bar{Q}_1Q_2Q_3)| \leqslant |\rho(Q_1\bar{Q}_2\bar{Q}_3)|$ and $|\rho(Q_1\bar{Q}_2Q_3)| > |\rho(\bar{Q}_1Q_2\bar{Q}_3)|$ and none of the sets are empty, then $\rho(Q_1\bar{Q}_2\bar{Q}_3)$ can be partitioned into two parts as follows:

$$\rho(Q_1\bar{Q}_2\bar{Q}_3) = \rho_1(Q_1\bar{Q}_2\bar{Q}_3) + \rho_2(Q_1\bar{Q}_2\bar{Q}_3),$$

where

$$|\rho_1(Q_1\bar{Q}_2\bar{Q}_3)| = |\rho(\bar{Q}_1Q_2Q_3)|$$

and $\rho(Q_1\bar{Q}_2Q_3)$ can be partitioned into two parts as follows:

$$\rho(Q_1\bar{Q}_2Q_3) = \rho_1(Q_1\bar{Q}_2Q_3) + \rho_2(Q_1\bar{Q}_2Q_3),$$

where

$$|\rho_1(Q_1\bar{Q}_2Q_3)| = |\rho(\bar{Q}_1Q_2\bar{Q}_3)|.$$

Then the following 2D organization is 2D C-R organization for $\{Q_1, Q_2, Q_3\}$.

$$\begin{array}{llllll} \rho(Q_1Q_2\bar{Q}_3), & \rho_1(Q_1\bar{Q}_2Q_3), & \rho_1(Q_1\bar{Q}_2\bar{Q}_3), & \rho(Q_1Q_2Q_3), & \rho_2(Q_1\bar{Q}_2Q_3), & \rho_2(Q_1\bar{Q}_2\bar{Q}_3) \\ & \rho(\bar{Q}_1Q_2\bar{Q}_3), & \rho(\bar{Q}_1Q_2Q_3), & & & , \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

If $\rho(Q_1Q_2\bar{Q}_3) = \varnothing$ or $\rho(Q_1Q_2Q_3) = \varnothing$ or $\rho(\bar{Q}_1\bar{Q}_2Q_3) = \varnothing$ or $\rho_1(Q_1\bar{Q}_2\bar{Q}_3) \cup \rho(\bar{Q}_1Q_2Q_3) = \varnothing$ or $\rho_1(Q_1\bar{Q}_2Q_3) \cup \rho(\bar{Q}_1Q_2\bar{Q}_3) = \varnothing$ the above 2D organization is still a 2D C-R organization.

If $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ then the 2D C-R organization is

$$\begin{array}{lllll} \rho(\bar{Q}_1Q_2\bar{Q}_3), & \rho(Q_1Q_2\bar{Q}_3), & \rho_1(Q_1\bar{Q}_2\bar{Q}_3), & \rho(Q_1Q_2Q_3), & \rho_2(Q_1\bar{Q}_2\bar{Q}_3) \\ & & \rho(\bar{Q}_1Q_2Q_3), & & , \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

If $\rho(\bar{Q}_1 Q_2 \bar{Q}_3) = \varnothing$ then the 2D C-R organization is

$$\begin{array}{l} \rho(Q_1 Q_2 \bar{Q}_3),\ \rho_1(Q_1 \bar{Q}_2 \bar{Q}_3),\ \rho(Q_1 Q_2 Q_3),\ \rho(Q_1 \bar{Q}_2 Q_3),\ \rho_2(Q_1 \bar{Q}_2 \bar{Q}_3) \\ \qquad\qquad \rho(\bar{Q}_1 Q_2 Q_3), \qquad\qquad\qquad\qquad\qquad , \rho(\bar{Q}_1 \bar{Q}_2 Q_3). \end{array}$$

If $\rho(\bar{Q}_1 Q_2 Q_3) = \varnothing$ then the 2D C-R organization is

$$\begin{array}{l} \rho(Q_1 Q_2 \bar{Q}_3),\ \rho_1(Q_1 \bar{Q}_2 Q_3),\ \rho(Q_1 Q_2 Q_3),\ \rho_2(Q_1 \bar{Q}_2 Q_3),\ \rho(Q_1 \bar{Q}_2 \bar{Q}_3) \\ \qquad\qquad \rho(\bar{Q}_1 Q_2 \bar{Q}_3), \qquad\qquad\qquad\qquad\qquad , \rho(\bar{Q}_1 \bar{Q}_2 Q_3) \end{array}$$

If $\rho(\bar{Q}_1 Q_2 \bar{Q}_3) = \varnothing$ and $\rho(\bar{Q}_1 Q_2 Q_3) = \varnothing$ then the 2D C-R organization is

$$\begin{array}{r} \rho(Q_1 Q_2 \bar{Q}_3),\ \rho(Q_1 Q_2 Q_3),\ \rho(Q_1 \bar{Q}_2 Q_3),\ \rho(Q_1 \bar{Q}_2 \bar{Q}_3) \\ \rho(\bar{Q}_1 \bar{Q}_2 Q_3) \end{array}$$

Case II. If $|\rho(Q_1 \bar{Q}_2 \bar{Q}_3)| < |\rho(\bar{Q}_1 Q_2 Q_3)|$ and $|\rho(Q_1 \bar{Q}_2 Q_3)| \leqslant |\rho(\bar{Q}_1 Q_2 \bar{Q}_3)|$ and none of the seven sets are empty than a 2D C-R organization can be constructed in the following manner:

Let

$$\rho(\bar{Q}_1 Q_2 Q_3) = \rho_1(\bar{Q}_1 Q_2 Q_3) + \rho_2(\bar{Q}_1 Q_2 Q_3),$$

where

$$|\rho_1(\bar{Q}_1 Q_2 Q_3)| = |\rho(Q_1 \bar{Q}_2 \bar{Q}_3)|,$$

and

$$\rho(\bar{Q}_1 Q_2 \bar{Q}_3) = \rho_1(\bar{Q}_1 Q_2 \bar{Q}_3) + \rho_2(\bar{Q}_1 Q_2 \bar{Q}_3),$$

where

$$|\rho_1(\bar{Q}_1 Q_2 \bar{Q}_3)| = |\rho(Q_1 \bar{Q}_2 Q_3)|$$

then a 2D C-R organization is as follows:

$$\begin{array}{l} \qquad\quad \rho(Q_1 Q_2 \bar{Q}_3),\ \ \rho(Q_1 \bar{Q}_2 Q_3),\ \ \rho(Q_1 \bar{Q}_2 \bar{Q}_3),\ \rho(Q_1 Q_2 Q_3), \\ (\bar{Q}_1 Q_2 \bar{Q}_3), \qquad\qquad , \rho_1(\bar{Q}_1 Q_2 \bar{Q}_3),\ \rho_1(\bar{Q}_1 Q_2 Q_3), \qquad\qquad , \rho_2(\bar{Q}_1 Q_2 Q_3),\ \rho(\bar{Q}_1 \bar{Q}_2 Q_3). \end{array}$$

It can be shown that the 2D C-R organization is preserved when the different sets are null, but the following three situations need some modifications.

If $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ then the 2D C-R organization is

$$\begin{array}{l} \qquad\qquad \rho(Q_1Q_2\bar{Q}_3),\ \rho(Q_1\bar{Q}_2\bar{Q}_3),\ \rho(Q_1Q_2Q_3), \\ \rho(\bar{Q}_1Q_2\bar{Q}_3), \qquad\qquad ,\ \rho_1(\bar{Q}_1Q_2Q_3), \qquad\qquad ,\ \rho_2(\bar{Q}_1Q_2Q_3),\ \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

If $\rho(Q_1\bar{Q}_2\bar{Q}_3) = \varnothing$ then a 2D C-R organization is

$$\begin{array}{l} \qquad\qquad \rho(Q_1Q_2\bar{Q}_3),\ \rho(Q_1\bar{Q}_2Q_3),\ \rho(Q_1Q_2Q_3), \\ \rho_2(\bar{Q}_1Q_2\bar{Q}_3), \qquad\qquad ,\ \rho_1(\bar{Q}_1Q_2\bar{Q}_3), \qquad\qquad ,\ \rho(\bar{Q}_1Q_2Q_3),\ \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

If $\rho(Q_1\bar{Q}_2\bar{Q}_3) = \varnothing$ and $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ then a 2D C-R organization is

$$\rho(\bar{Q}_1Q_2Q_3),\ \rho(Q_1Q_2\bar{Q}_3),\ \rho(Q_1Q_2Q_3),\ \rho(\bar{Q}_1Q_2Q_3),\ \rho(\bar{Q}_1\bar{Q}_2Q_3)$$

In the case when $\rho(\bar{Q}_1Q_2\bar{Q}_3) = \varnothing$ it implies that $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ this fact is needed to show that a 2D C-R organization exists when $\rho(\bar{Q}_1Q_2\bar{Q}_3) = \varnothing$.

Case III. If $|\rho(\bar{Q}_1Q_2Q_3)| \leqslant |\rho(Q_1\bar{Q}_2\bar{Q}_3)|$ and $|\rho(Q_1\bar{Q}_2Q_3)| \leqslant |\rho(\bar{Q}_1Q_2\bar{Q}_3)|$ and the sets are nonempty then a 2D C-R organization can be constructed as follows:
Let

$$\rho(\bar{Q}_1Q_2\bar{Q}_3) = \rho_1(\bar{Q}_1Q_2\bar{Q}_3) + \rho_2(\bar{Q}_1Q_2\bar{Q}_3),$$

where

$$|\rho_1(\bar{Q}_1Q_2\bar{Q}_3)| = |\rho(Q_1\bar{Q}_2Q_3)|$$

and

$$\rho(Q_1\bar{Q}_2\bar{Q}_3) = \rho_1(Q_1\bar{Q}_2\bar{Q}_3) + \rho_2(Q_1\bar{Q}_2\bar{Q}_3),$$

where

$$|\rho_1(Q_1\bar{Q}_2\bar{Q}_3)| = |\rho(\bar{Q}_1Q_2Q_3)|$$

then a 2D C-R organization is as follows:

$$\begin{array}{r} \rho_2(\bar{Q}_1Q_2\bar{Q}_3),\ \rho(Q_1Q_2\bar{Q}_3),\ \rho(Q_1\bar{Q}_2Q_3),\ \rho_1(Q_1\bar{Q}_2\bar{Q}_3),\ \rho(Q_1Q_2Q_3),\ \rho_2(Q_1\bar{Q}_2\bar{Q}_3) \\ \rho_1(\bar{Q}_1Q_2\bar{Q}_3),\ \rho(\bar{Q}_1Q_2Q_3), \qquad\qquad ,\ \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

It is easy to show that when the different sets are null sets the 2D C-R organization still holds but the following situations need special attention.
When $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ the 2D C-R organization is

$$\begin{array}{r} \rho(\bar{Q}_1Q_2\bar{Q}_3),\ \rho(Q_1Q_2\bar{Q}_3),\ \rho_1(Q_1\bar{Q}_2\bar{Q}_3),\ \rho(Q_1Q_2Q_3),\ \rho_2(Q_1\bar{Q}_2\bar{Q}_3) \\ \rho(\bar{Q}_1Q_2Q_3), \qquad\qquad ,\ \rho(\bar{Q}_1\bar{Q}_2Q_3). \end{array}$$

When $\rho(\bar{Q}_1Q_2Q_3) = \varnothing$ then a 2D C-R organization is

$$\begin{array}{l} \rho_2(\bar{Q}_1Q_2\bar{Q}_3), \rho(Q_1Q_2\bar{Q}_3), \rho(Q_1\bar{Q}_2Q_3), \rho(Q_1Q_2Q_3), \rho(Q_1\bar{Q}_2\bar{Q}_3) \\ \qquad\qquad \rho_1(\bar{Q}_1Q_2\bar{Q}_3), \qquad\qquad , \rho(\bar{Q}_1Q_2Q_3) \end{array}$$

when $\rho(\bar{Q}_1Q_2Q_3) = \varnothing$ and $\rho(Q_1\bar{Q}_2Q_3) = \varnothing$ then a 2D C-R organization is

$$\begin{array}{l} \rho(\bar{Q}_1Q_2\bar{Q}_3), \rho(Q_1Q_2\bar{Q}_3), \rho(Q_1Q_2Q_3), \rho(Q_1\bar{Q}_2\bar{Q}_3) \\ \qquad\qquad\qquad\qquad \rho(\bar{Q}_1\bar{Q}_2Q_3) \end{array}$$

This completes the proof of Lemma 4.1.

LEMMA 4.2. *A sufficient condition for the existence of the 2D C-R property for $\{Q_1, Q_2, Q_3\}$ is the negation of either or both of the following conditions*:

$$|\rho(Q_1Q_2\bar{Q}_3)| > |\rho(\bar{Q}_1\bar{Q}_2Q_3)|; \ |\rho(Q_1\bar{Q}_2Q_3)| > |\rho(\bar{Q}_1Q_2\bar{Q}_3)|$$

Proof of Lemma 4.2. The proof of this lemma is exactly similar to that of Lemma 4.1. The basic 2D C-R organization is as follows:

$$\begin{array}{l} \rho(Q_1\bar{Q}_2\bar{Q}_3), \rho(Q_1Q_2\bar{Q}_3), \rho(Q_1\bar{Q}_2Q_3), \rho(Q_1Q_2Q_3), \rho(\bar{Q}_1Q_2Q_3) \\ \qquad \rho(\bar{Q}_1\bar{Q}_2Q_3), \rho(\bar{Q}_1Q_2\bar{Q}_3). \end{array}$$

LEMMA 4.3. *A sufficient condition for the existence of the 2D C-R property for $\{Q_1, Q_2, Q_3\}$ is the negation of either or both of the following conditions*:

$$|\rho(Q_1Q_2\bar{Q}_3)| > |\rho(\bar{Q}_1\bar{Q}_2Q_3)|; \ |\rho(\bar{Q}_1Q_2Q_3)| > |\rho(Q_1\bar{Q}_2\bar{Q}_3)|$$

Proof of Lemma 4.3. The proof of this lemma is exactly similar to that of Lemma 4.1. The basic 2D C-R organization is as follows:

$$\begin{array}{l} \rho(Q_1\bar{Q}_2Q_3), \rho(Q_1Q_2Q_3), \rho(Q_1Q_2\bar{Q}_3), \rho(\bar{Q}_1Q_2Q_3), \rho(\bar{Q}_1Q_2\bar{Q}_3) \\ \qquad\qquad \rho(\bar{Q}_1\bar{Q}_2Q_3), \rho(Q_1\bar{Q}_2\bar{Q}_3) \end{array}$$

The proof of Theorem 4 is obtained by combining the results of Lemmas 4.1, 4.2, and 4.3.

Proof of Theorem 6. Suppose $\{Q\}$ is a query set and $\{R\}$ is a record set. $\{R\}$ is partitioned into subsets $\{r_1\}, \{r_2\},..., \{r_n\}$ and their incidence domains w.r.t. $\{Q\}$ is given by

$$\rho^{-1}(\{r_1\}), \rho^{-1}(\{r_2\}),..., \rho^{-1}(\{r_n\}).$$

The incidence domain of each $\{r_i\}$, $i = 1, 2,..., n$ is disjoint w.r.t. $\{Q\}$.

Case 1. Suppose

$$\rho^{-1}(\{r_1\}) \supseteq \rho^{-1}(\{r_2\}) \supseteq \rho^{-1}(\{r_3\}) \supseteq \cdots \supseteq \rho^{-1}(\{r_n\}).$$

A 2D organization can be formed out of these subsets in the following manner

$$\{r_1\}, \{r_2\}, \ldots, \{r_n\},$$

where the $\{r_i\}$'s are secondary arrays of the 2D organization. In this organization if $Q_i \in \{Q\}$ is not relevant to any record in $\{r_j\}$ then it is not relevant to any record in $\{r_{j+1}\}, \ldots, \{r_n\}$, and if a record in $\{r_{j-1}\}$ is pertinent to Q_i then in each of the secondary arrays $\{r_{j-1}\}, \{r_{j-2}\}, \ldots, \{r_1\}$ there exists one and only one record relevant to Q_i. Thus, $\{Q\}$ has the 2D C-R property w.r.t. $\{R\}$.

Case II. If $\rho^{-1}(\{r_1\}) \subseteq \rho^{-1}(\{r_2\}) \subseteq \cdots \subseteq \rho^{-1}(\{r_n\})$ then the proof is exactly similar to that of Case I.

Case III. Suppose

$$\rho^{-1}(\{r_1\}) \subseteq \rho^{-1}(\{r_2\}) \subseteq \cdots \subseteq \rho^{-1}(\{r_I\}) \supseteq \rho^{-1}(\{r_{I+1}\}) \supseteq \cdots \supseteq \rho^{-1}(\{r_n\}).$$

If $Q_i \in \{Q\}$ is not relevant to any record in $\{r_j\}$ and $j < I$ then it is not relevant to any record in $\{r_{j-1}\}, \{r_{j-2}\}, \ldots, \{r_1\}$. If Q_i is relevant to a record in $\{r_{j+1}\}$ for $j > I$ then a record pertinent to Q_i may exist in $\{r_{j+2}\}, \{r_{j+3}\}$ and so on up to $\{r_n\}$. If a record pertinent to Q_i does not exist in $\{r_J\}$ for some J which is greater than $j + 1$, then no record pertinent to Q_i exists in $\{r_{J+1}\}, \{r_{J+2}\}, \ldots, \{r_n\}$. Hence, the records pertinent to Q_i are stored in a 2D C-R organization.

This completes the proof.

Proof of Theorem 7. Suppose the query set $\{Q\}$ has the 2D C-R property w.r.t. $\{R\}$ and a 2D C-R organization is $\{r_1\}, \{r_2\}, \ldots, \{r_n\}$, where $\{r_i\}$, $i = 1, 2, \ldots, n$ are the secondary arrays of the 2D C-R organization. Suppose $\{R_1\}$ is another set of records which is disjoint from $\{R\}$ and the records in $\{R_1\}$ have disjoint incidence domain which are subsets of $\{Q\}$, hence, $\{R_1\}$ is qualified to be a secondary array in the 2D C-R organization.

Let

$$\rho^{-1}(\{r_i\}) \cap \rho^{-1}(\{r_{i+1}\}) \subseteq \bigcup_{R_j \in \{R_1\}} \rho^{-1}(R_j) \subseteq \rho^{-1}(\{r_i\}) \cup \rho^{-1}(\{r_{i+1}\})$$

for some i.

Since $\rho^{-1}(\{r_i\}) \cap \rho^{-1}(\{r_{i+1}\}) \subseteq \bigcup_{R_j \in \{R_1\}} \rho^{-1}(R_j)$, if $\{R_1\}$ is inserted as a secondary array between $\{r_i\}$ and $\{r_{i+1}\}$ then the retrieval paths of all queries which travel from $\{r_i\}$ to $\{r_{i+1}\}$ will remain uninterrupted.

Since $\rho^{-1}(\{r_i\}) \cup \rho^{-1}(\{r_{i+1}\}) \supseteq \bigcup_{R_j \in \{R_1\}} \rho^{-1}(R_j)$ if $\{R_1\}$ is inserted as a secondary array between $\{r_i\}$ and $\{r_{i+1}\}$ then the retrieval paths of some queries which terminate (or start) at $\{r_i\}$ will now terminate (or start) at $\{R_1\}$. The retrieval paths of some queries which start (or terminate) at $\{r_{i+1}\}$ will now start (or terminate) at $\{R_1\}$. Those queries which are not included in the incidence domain of $\{R_1\}$ will not be affected.

This completes the proof.

Acknowledgments

The author would like to thank Dr. M. E. Senko, Dr. R. G. Casey, and Dr. E. B. Altman of Information Sciences Department of IBM Research Laboratory in San Jose for their valuable comments and suggestions in the paper.

The author would like to thank the referee for some helpful suggestions.

Received: July 8, 1972; revised: December, 1972

References

1. Ghosh, S. P. (1970), On the theory of consecutive storage of relevant records, IBM Research Report No. RJ 708 and *J. Inform. Sci.* **6**, 1–9 (1973).
2. Ghosh, S. P. (1972), File organization: The consecutive retrieval property, *Comm. ACM* **15**, 802–808. Also published as IBM Research Report No. RJ 765 (1970).